

End-to-end learning

speech recognition example:

→ audio $\xrightarrow{\text{MFCC}}$ features $\xrightarrow{\text{ML}}$ phonemes $\xrightarrow{\quad}$ words $\xrightarrow{\quad}$ transcripts

End-to-end:

→ audio $\xrightarrow{\quad}$ transcripts

Pros: let the data speak

less hand-designing of components needed

cons: may need large amount of data

Excludes potentially useful hand-designed components.

✗ Applying end-to-end deep learning:

Key question: Do you have sufficient data to learn a function of the complexity needed to map x to y ?